

TechnoFeature™

Practice management and technology articles written by experts.

eDiscovery: Everything You Need to Know About Analysis and Production

By Bruce A. Olson

INTRODUCTION

Once the process of identification, collection and processing, and review of relevant electronically stored information (ESI) has been accomplished, it is time to move on to the analysis and production phase.

ANALYSIS

There are a number of technology-based solutions that will help you analyze your collection of electronic materials. They include the use of search features, derived metadata, guided navigation, and visualization.

Search

Most people who use litigation support technology are familiar with basic searching within a given database of materials. Typically, objective information about each item is coded during the processing phase, and additional subjective coding is done during the review process. The coded information exists in identifiable fields that comprise a searchable database. You can analyze a specific field category, drill down using word searches within that field, or do word searches across multiple fields.

Most programs will enable you to use fuzzy searches or wildcards with given words, or you can broaden the search to include identifiable phrases. Proximity searching is also available. Typically, you can refine your search by date ranges.

As your search skills improve you will be able to develop more complex Boolean queries to further refine your results. Once the dataset has been identified it can be sorted, tallied, and ranked for relevance, and ultimately tagged for production.

Derived Metadata

A more advanced way to search is to use meta-data and derived metadata. With the advent of eDiscovery one must always consider whether the requested information should be produced in native format or in some other format that provides some level of capture of the associated metadata. This information can be added to the database during the loading process, and can be used to automatically populate certain fields.

With the advent of eDiscovery one must always consider if the requested information should be produced in native format.

With some more advanced programs, this concept can be expanded so that key phrases, contextual groups, directory information and categories, and even clusters can automatically be identified and used to filter a subset of specific materials that can speed your analysis. Issue coding or key phrase coding can be also performed during the review process. This will enable you to easily access specific categories of documents.

The key to success in this area is to ensure that all reviewers are categorizing the information in the same way. Automatic lookup tables, clear and concise instructions in coding manuals, and ongoing supervision of data entry can be used to ensure the success of this approach.

(Continued on next page)

Guided Navigation

In some cases searching by contextual groups can be a powerful analytic tool. Email threads, including the capture of attachments to emails, are a common example of a contextual group. Searching by directory information such as names, email addresses, and department groupings can result in the production of a dataset that helps you analyze the interrelationships among a given group of people as it relates to a particular document or set of documents.

Finally, category and cluster searching can be performed. Category searches include search by author, date created, date modified, etc. Clustering could include email and attachments, or clusters of document types that are related such as all design drawings related to a particular product or all contract documents related to a given project or transaction.

Take care when creating your review team in order to obtain the best use of the knowledge base and legal experience of the individual reviewers.

This concept becomes particularly useful when combined with programs that offer visualization tools to help you determine what areas to analyze first through the use of topical and sub-topical clustering representations.

Visualization

Visualization tools are typically created using either a social network analysis, a context group structure analysis, or a topical cluster analysis. Social networking analysis maps the relationships between people, groups, and organizations and visually demonstrates the flow of information between these groups. Context group analysis looks at a context group such as an email thread, and maps the flow of information between the senders and recipients

in a visual fashion. Cluster analysis visually identifies relationships typically based on keywords or phrases.

For example, there may be a context group of Smith contract documents and Jones contract documents, with approximately twice as many documents in the Smith category. If you want to analyze contract documents, you would visually note that there are many more dots within a circle representing the Smith documents as compared to the Jones documents. Based on the volume of materials identified visually, you may decide that your analysis should start with the larger group. This same type of analysis can also be done using bar graphs that show more hits for a given word or a given date, which helps you focus your initial attention on the area that seems to have the greatest activity.

As an aside, screen captures of these visualization tools that are placed in some type of presentation program, followed by links to images of the specific documents, can be a powerful demonstrative tool at trial or in mediation. They can set the tone and scope of what's important for the fact finder on a more global basis before you drill down to the specific evidence.

In my opinion, good trial attorneys will constantly be on the alert for new and innovative ways to present information to a jury as they use the new and developing tools to analyze electronically stored information.

Tips

A good strategy of analysis using the available tools to identify and focus on relevant subsets of information can speed the process of understanding the components of a particular case. Once again, if the attorney does not know how to use the tools, whether on a hands-on basis or at least on a theoretical basis, the ability to rapidly address what is or is not relevant to your case will be severely compromised. Failure to use these techniques can result in the production of too much information, the production of irrelevant information, and possibly the inadvertent production of privileged or confidential information.

(Continued on next page)

Take care when creating your review team in order to obtain the highest and best use of the knowledge base and legal experience of the individual reviewers. Assigning identified subsets of information to specific qualified individuals will also result in a consistent analytical approach.

PRODUCTION

Having completed your analysis you should now have determined what you will produce to the opposing side.

Claw Back Agreements

Keep in mind that no matter how sophisticated your technological skills, tools, and methods are, given the volume of information that typically needs to be produced in an electronic production the risk of inadvertent disclosure is high. Therefore, if you haven't considered it before, be sure to obtain an appropriate claw back agreement before you turn over your information. Ideally this issue will be addressed in the initial meet and confer.

Format

In addition to claw back provisions, you should have addressed the methods and form of production. Under Rule 34 the requesting party may specify the desired form or forms of production for identified categories of information, and it is good practice to always do so. If they fail to make such a request the information must still be produced in a form or forms in which it is ordinarily maintained or in a form or forms that are reasonably usable.

The latter statement gives you some latitude in the form of production, such as exporting information from a proprietary database into an Access database so the resulting file becomes usable by all parties, but the form of production must still be electronic. Gone are the days (absent agreement) of TIFFing or PDFing the information and simply burning it to disk and giving it to the other side. Abusive strategies, actually reported in case law, such as printing all the electronic information to paper, manually Bates stamping it, scanning it as TIFFs and then turning it over on disk and claiming it is an electronic production, will simply not be countenanced.

(Having said that, you should be aware that you do have the right under the amended rules to agree to forgo an electronic production. Some are still so wedded to paper that both sides would prefer to remain in a paper world or to forgo electronic discovery altogether.)

Often times you will be dealing with mixed nodes of production, where paper and ESI are both called for. In those circumstances, it makes sense to image the paper documents. If you are going to do that you should attempt to agree with the other side on the image format, typically TIFF or PDF; whether the images should be OCR'd during the scanning process; what type of load file to request; and the delivery medium, e.g., CDs, DVDs, flash drives, external hard drives, etc.

Everyone should use the available technologies to analyze electronically stored information to identify discoverable and producible information.

You also have the right to agree to ignore the collection and production of metadata. Keep in mind, however, that once you've made that decision, it is very difficult to go back and obtain an order from the court forcing the other side to make a new production that includes the metadata. Any waiver of metadata should be a knowing waiver and you must be prepared to suffer the consequences.

Other Considerations

The process of producing electronically stored information can be time consuming and you should attempt to negotiate reasonable production schedules.

You should also consider whether it is appropriate in a given case to use a third-party vendor to host the data and make it available via the Internet. While there are charges involved in outsourcing the data, it also eliminates the IT overhead expense

(Continued on next page)

associated with in-house hosting. The cost can be shared among multiple parties. If you use the vendors search tools it also provides a certain level of commonality of approach in working with the information, which may make it easier to collaborate.

CONCLUSION

In the end, everyone should use the available technologies to analyze electronically stored information to identify discoverable and producible information. It is only through the use of such technology that one can successfully and cost-effectively manage the large volumes of data that exists in our electronic world.

Copyright 2007 Bruce A. Olson. All rights reserved.

ABOUT THE AUTHOR

Bruce A. Olson, is a shareholder in the Milwaukee based law firm of [Davis & Kuelthau, S.C.](#) A trial attorney and nationally recognized legal technologist

focusing primarily on the areas of electronic discovery and litigation technology, Olson is AV rated and Board Certified by the National Board of Trial Advocacy. He is co-author of *The Electronic Evidence and Discovery Handbook: Forms, Checklists and Guidelines*, published by the American Bar Association. He received the prestigious "TechnoLawyer of the Year 2002" Award from the TechnoLawyer Community, and was Chair of ABA TECHSHOW 2004, Vice Chair of ABA TECHSHOW 2003, and served on the TECHSHOW Board of Directors from 2000-2004.

Contact Bruce:

E: bolson@dkattorneys.com

T: (920) 431-2230

About TechnoFeature

Published on Tuesdays, *TechnoFeature* is a weekly newsletter containing in-depth articles written by leading legal technology and practice management experts, many of whom have become "household names" in the legal profession. Most of these articles are TechnoLawyer exclusives, but we also scour regional legal publications for superb articles that you probably missed the first time around.